

Introductory event for ESRC Research Methods Programme Buxton 19-20 May 2003

Assessing quality in grant applications: Reflections from experience on the ESRC Grants Board

Hazel Genn¹

1. Introduction

The background paper prepared for this meeting raises a number of important issues on the assessment of research “quality”. In my analysis of the process by which those assessments have been made by the ESRC Grants Board I will also comment on two issues raised in the background paper. First, the broad and important question of whether disciplinary differences interfere with judgements about quality in grant applications, and, in passing, the problem of a interpreting referees’ assessments in the decision-making process.

2. Understanding the Grants Board assessment process

To put my reflections in context I want to start by outlining the assessment process used by the Grants Board. This is important in order to understand the various inputs into decision-making and how comparisons across disciplines are made.

Each application to the Board is assigned to a Board Assessor and Discussant. The allocation is made on a disciplinary basis. In this way applications receive specific attention from two Board members with broad disciplinary competence. Applications are then sent to several referees comprising at least one nominated referee and two or more referees selected by the office from a referee database.

Once the referees’ comments have been received, the Board Assessor and Board Discussant make an independent judgement on the proposal and assign a Grade A, B or R. Within the Alpha grades applications are ranked from 1 to 5 based on priority for funding. A grade of A1 represents a project of the highest scientific quality and highest priority for funding.

Prior to the Board meeting the Assessor and Discussant grades are combined to produce an index. All Board members receive copies of all proposals, referees comments, and Assessors’ and Discussants’ comments. At Board meetings the Assessors and Discussants present their views to the whole Board and other members have the opportunity to discuss each proposal before a decision about funding is made.

¹ Professor of Socio-Legal Studies, Faculty of Laws, University College London. The views expressed here are personal, based on my perception of the process. Others might have a different interpretation. They also reflect my own epistemological orientation and approach to research which, again, will certainly differ from that of others. The thoughts are therefore offered as a basis for discussion. I am grateful to Sharon Witherspoon, Deputy Director of the Nuffield Foundation with whom I have discussed the issues raised in this paper.

Experience on the Board therefore gives considerable opportunity to appreciate disciplinary differences and similarities in the criteria central to judgements of scientific “quality”.

3. Dealing with referees comments

Board assessors make their own independent judgement on the quality of a proposal. This judgement is *informed* by the views of the referees, but the views of referees cannot be, and are not, determinative. Why is this so?

First, referee responses are very variable. Some are thoughtful, careful, and well-reasoned. Some are short and effusive - some are short and dismissive. Sometimes there is a poor fit between grade and reasoning and this can go in either direction. An assessor can be faced with a very supportive commentary with no significant criticism accompanied by a grade of A-. Alternatively they may have to deal with a Grade of A+ accompanied by a commentary that raises significant and fundamental concerns about some aspect of design. **It is my perception that there may be some disciplinary differences in this respect i.e. that some disciplines seem to be more reluctant than others to award A+.**

Second, there may be a difference of view among the three or four referees commenting on a proposal. The various judgements of referees are not always consistent (contrary to the suggestion in the background paper to this meeting). This sometimes happens when a proposal is situated at the interface between different disciplines and responses to the proposal vary between disciplines. On the other hand, one or two referees may identify a particular problem with the proposal that is not picked up by other referees on the same proposal and the result is a split in grades.

For these reasons referees comments and grades often have to be *interpreted* by the assessors and discussants and then, possibly, by the Board as a whole. In this process of interpretation it is possible that an application with high grades may not be funded. A proposal with consistently low grades from referees would not be funded.

4. Unbundling The Judgement Of Scientific Quality

What then are the elements that are considered in reaching a judgement on applications and how do they relate to “scientific quality”?

Subject matter of the research

Issues about subject matter that will be considered in reaching funding decisions may relate to the question of scientific quality, to priority for funding, or to both.

The originality of the proposed research

This issue *does* relate to judgements about scientific quality since research that essentially duplicates what has gone before without any potential increase in knowledge or understanding is unlikely to be regarded as high quality. Proposals can promise originality, for example, in the broad topic of the research, in taking a novel methodological approach to a particular subject, or in the use of innovative methods.

Inherent interest and Timeliness of the subject

These issues relate more to questions about priority for funding rather than the quality of the science. Good quality science may not achieve a high priority for funding if the subject is not felt to be particularly interesting or of sufficient importance to merit funding in competition with other good quality applications. There is never enough money available to fund all applications of scientific value.

Research Design

Absolutely central to judgements about scientific quality, in my view, is the following question: **Is the research design likely to produce credible evidence that bears directly on the questions being asked so as to achieve the research objectives?**

Applicants fail this test for various reasons but the most common are inadequate information on design and methods; poor-fit between proposed methods and research questions; inappropriate approach to analysis or absent information on proposed analysis.

Specification of research objectives and research questions.

In order to test the fit between research design and questions it is necessary to have a relatively clear statement of the broad objectives of the research and the specific questions that the research will address. The specification of objectives and research questions in an application can be made with a greater or lesser degree of precision and this is an area where there is understandable and acceptable disciplinary variation. The hypotheses to be tested in experimental research are likely to be elaborately specified, whereas the questions underlying a period of participant observation may be much more fluid, and it is reasonable to expect that they will evolve and change during the period of observation. While these disciplinary differences can be and are accommodated within the Board's decision-making process, there must be some ascertainable research problem or questions identified. Without a reasonably precise set of issues and articulation of questions at the heart of the proposal it is difficult to make reasoned judgements about quality, unless one is simply going on feel or backing a "good person" - which the Board is understandably reluctant to do.

Insufficient information on research design

A common difficulty in making judgements about scientific quality is simply the paucity of information provided by applicants about their research design. There are a number of possible causes of this problem. Applicants may not appreciate the centrality of research design in decision-making. Many applicants exhaust themselves establishing the context for the research, reviewing work already done, and mounting a comprehensive argument for the significance of the research problem. They then omit to articulate the questions and explain the design in sufficient detail for a judgement to be made. Others may lack confidence in their proposed approach and wrongly hope that providing minimal information will give fewer grounds for criticism. In some cases the absence of information may simply reflect the applicant's failure to think through the design in advance on the assumption that this is something that can be done once the project has been funded. Without adequate information about how the research objectives are to

be accomplished and whether the choice of approach is appropriate, it is difficult to make a decision on the quality of the science. In the days when applications could be graded as Revise and Resubmit a promising subject that was deficient in its specification of methods could possibly be saved through this route. Without that option, it is possible for the Board to make an award conditional on receiving a more detailed specification, but this would only occur if the amount of detail required were relatively minor. Where there are substantial unanswered questions about the methods, a project is likely to be marked down on quality even though the subject-matter of the research appears highly original, interesting and timely. With intense competition for funding these are some of the matters that will distinguish the good from the best.

Fit between design and research questions

The question of fit is central to judgements about quality. Are the proposed methods appropriate to the research objectives and questions? This question of fit is not a narrow technical issue. It raises fundamental questions about researchers' understandings of how to conduct research at the most basic level. It raises questions about understanding of the potential and limitations of different research methods; about what types of analysis are appropriate to different types of research data; about whether and what level of generalisation can be supported by what kind of data; and even about what kind of evidence will have a bearing on research questions. It goes to the heart of judgements about quality and raises issues about how social researchers are trained, what they understand by research, and how social research justifies its claims to be taken seriously.

The most common and egregious example of poor fit is where the researcher specifies objectives and research questions that require large-scale quantitative evidence, but then proposes to use small-scale qualitative research techniques. This isn't a problem of a sample that is going to be too small - it represents a fundamental misunderstanding of qualitative methodology.

A different example of poor fit is when the applicant specifies a quantitative approach but the sample is too small to accomplish the proposed analysis. This may occur through a lack of understanding of the potential and limitations of quantitative methods and failure to think through the type of analysis needed in order to answer the specified questions.

Another version of mismatch is where the applicant is enthusiastic about adopting a multi-method approach. A design may be proposed that includes a bit of everything, but in which no one element is likely to yield data of either sufficient quantity, depth or quality to provide credible evidence bearing on the research questions.

The failure to understand the appropriateness of different methodologies for different types of research question is a problem worthy of detailed consideration by a meeting of this kind. Is it a manifestation of the dislocation of data collection techniques from their philosophical underpinning? Does it raise questions about how research methods are taught in universities and are there clear disciplinary differences here that are worth exploring? Is there a widespread lack of competence, confidence or interest in quantitative methods despite the fact that researchers identify and are keen to address questions that require quantitative evidence? These questions, of course, raise but do not answer another of the issues discussed in the background paper to the meeting – how to establish an “acceptable” quality of evidence for the questions posed?

5. Fitness-for-purpose as a generic test

Assessing the proposed research design in relation to the research objectives and specific research questions is a process of judgement that is *not* distorted by different disciplinary research traditions. Indeed, making a core judgement in this focused way helps to avoid that kind of misunderstanding and distortion. There are clear differences in approach between anthropologists, experimental psychologists, socio-linguists, management scientists, macro-economists etc. In my view it is too easy for a discipline to criticise judgements of scientific quality on the ground that “no one else understands what we do.” Different disciplines do have different traditions in their approach to inquiry. An economist may not envisage using the qualitative techniques of an anthropological field researcher – but then she would be unlikely to have the same kinds of research objectives or questions as the anthropologist has in mind when she begins an extensive period of observation in the field. On the other hand, while the anthropologist will not frame research questions in terms of hypotheses to be tested, it is reasonable to expect some degree of precision in terms of the issues exciting her interest and the questions that will guide observation.

In a multi-disciplinary decision-making body, questions of quality are therefore answered *not* by reference to a hierarchy of data collection methods, but through a fitness-for-purpose test that recognises a wide range of purposes and a similarly wide range of legitimate approaches in social science research.